Advances and Applications in Statistics
© 2026 Pushpa Publishing House, Prayagraj, India
https://pphmjopenaccess.com
https://doi.org/10.17654/0972361726019
Volume 93, Number 3, 2026, Pages 361-382
P-ISSN: 0972-3617
E-ISSN: 3049-2165

EFFICIENCY-ENHANCED POPULATION MEAN ESTIMATION UNDER STRATIFIED SAMPLING WITH INDETERMINATE DATA

Muhammad Waqar Hussain, Maria Javed* and Muhammad Irfan

Department of Statistics
Government College University
Faisalabad, Pakistan
e-mail: mariajaved@gcuf.edu.pk

Abstract

Neutrosophic estimation is a major development in sampling theory that successfully tackles the problems posed by uncertain, indeterminate and unreliable data. This study estimates the population mean of the study variable by incorporating auxiliary information within a neutrosophic environment under stratified sampling. A neutrosophic stratified exponential ratio-type estimator is aimed at improving estimation accuracy in heterogeneous neutrosophic populations. The proposed estimator enhances both precision and reliability. It assesses the bias and mean square error (MSE) through

Received: November 19, 2025; Revised: January 1, 2026; Accepted: January 30, 2026
2020 Mathematics Subject Classification: 62D05.
Keywords and phrases: auxiliary information, mean squared error, neutrosophic data, percentage relative efficiency.

*Corresponding author

How to cite this article: Muhammad Waqar Hussain, Maria Javed and Muhammad Irfan, Efficiency-enhanced population mean estimation under stratified sampling with indeterminate data, Advances and Applications in Statistics 93(3) (2026), 361-382.
https://doi.org/10.17654/0972361726019

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Published Online: March 19, 2026

first-order approximations. The theoretical findings are supported by empirical evidence, which demonstrates the estimator's superior performance, achieving lower MSEs and higher percent relative efficiencies (PREs) when compared to conventional estimators. These results affirm the proposed estimator as a reliable and efficient tool for future applications in neutrosophic stratified sampling (NSS), particularly in environments characterized by indeterminate data.

1. Introduction

In sample surveys, the primary objective of researchers is to determine the most effective methods for estimating population parameters. This involves optimizing time, minimizing costs and reducing sampling errors. One key strategy for achieving these goals is the use of additional/supplementary information, which helps decrease sampling variability and improve the efficiency of estimators. This supplementary information often consists of auxiliary variables that are strongly correlated with the study variables. These variables are typically easy to access for all units under investigation or can be sourced from previous studies or historical data.

When a population is divided into subgroups characterized by minimal within-group variation and maximal between-group variation, simple random sampling fails to enhance the efficiency of estimator(s) or reduce sampling error. In these situations, stratified random sampling is more effective. A variety of estimators have been proposed within the stratified sampling framework, each offering specific advantages as well as potential limitations depending on the nature of the data and the study objectives. For reference, see Kadilar and Cingi [2], Koyuncu and Kadilar [3], Shabbir and Gupta [6], Ahmadini et al. [1] and Sharma and Malik [7]. Although these studies have made significant contributions to the advancement of stratified sampling, their methodologies are entirely based on exact and well-defined data. However, in real-world situations, data often contain uncertainties, vagueness, or imprecise information, which is commonly expressed in the form of intervals.

To address this limitation, neutrosophic statistics, a branch derived from fuzzy logic, has been developed to handle both certain and indeterminate aspects of uncertain data. Common examples include interval-based temperature measurements, regional variations in water levels, and dimensional changes in machinery resulting from measurement inaccuracies.

In recent years, researchers have increasingly focused on developing enhanced neutrosophic estimators for unknown population parameters under various sampling schemes and scenarios involving indeterminate data. For reference, see Thanh et al. [11], Rao and Aslam [5], Tahir et al. [10] and Kumar et al. [4]. Singh et al. [8] suggested neutrosophic separate ratio and regression estimators for stratified sampling, while Singh et al. [9] introduced product and ratio exponential estimators within a neutrosophic environment under stratified sampling. More recently, Ullah et al. [12] applied neutrosophic statistics to address the compromise optimum allocation problem in stratified sampling under uncertainty.

Research gap

Despite extensive advancements in stratified sampling and the recent addition of neutrosophic statistics to address data uncertainty, most existing estimators are either developed under the assumption of exact data or are limited to specific estimators and sampling schemes. Consequently, a noticeable gap exists in the literature regarding the development of more efficiency-enhanced neutrosophic estimator for the estimation of population mean under stratified sampling. Such an estimator should aim to (i) effectively utilize auxiliary information in the presence of interval-valued or indeterminate data, (ii) achieve improved efficiency compared to existing neutrosophic estimators, and (iii) validate the theoretical findings through empirical studies based on diverse real-life neutrosophic data sets.

2. Neutrosophic Stratified Sampling

Neutrosophic stratified random sampling (NSRS) is a powerful way that blends the structure of stratified sampling with the flexibility of neutrosophic

set theory. It ensures that all subgroups in a population are represented while effectively managing uncertainty, vagueness and indeterminacy in the data. Furthermore, this approach partitions the entire neutrosophic heterogeneous population into clearly defined neutrosophic homogeneous subgroups, similar to the structure of conventional stratified sampling. Within each subgroup or stratum, the units exhibit uniform characteristics, while substantial differences exist between the strata in terms of the study variables. Once the population is stratified, separate samples are drawn from each stratum using the neutrosophic simple random sampling technique.

Let N_N be the finite neutrosophic heterogeneous population and divided into neutrosophic homogeneous L strata, each of size N_{hN}, where $N_{hN} \in [N_{hL}, N_{hU}]$ such that $\sum_{h=1}^{L} N_{hN} = N_N$, $h = 1, 2, 3, ..., L$. After dividing the population into strata, we apply the neutrosophic simple random sampling method to each stratum and select distinct samples of size n_{hN}, where $n_{hN} \in [n_{hL}, n_{hU}]$ such that $\sum_{h=1}^{L} n_{hN} = n_N$. Consider Y_N and X_N be the neutrosophic study and auxiliary variable respectively. Let Y_{hjN} and X_{hjN} be the population values of the study and auxiliary variable respectively of the j^{th} unit $(j = 1, 2, ..., N_{hN})$ in the h^{th} stratum. Similarly, y_{hjN} and x_{hjN} be the pair of sample values. Some important notations under neutrosophic environment are listed in Table 1.

Table 1. Useful notations under neutrosophic environment

Description	Neutrosophic study variable	Neutrosophic auxiliary variable
Population mean	$\bar{Y}_N = \sum_{h=1}^{L} W_{hN} \bar{Y}_{hN}$, $\bar{Y}_N \in [\bar{Y}_L, \bar{Y}_U]$	$\bar{X}_N = \sum_{h=1}^{L} W_{hN} \bar{X}_{hN}$, $\bar{X}_N \in [\bar{X}_L, \bar{X}_U]$
Population mean of h^{th} stratum	$\bar{Y}_{hN} = \frac{1}{N_{hN}} \sum_{j=1}^{N_{hN}} Y_{hjN}$, $\bar{Y}_{hN} \in [\bar{Y}_{hL}, \bar{Y}_{hU}]$	$\bar{X}_{hN} = \frac{1}{N_{hN}} \sum_{j=1}^{N_{hN}} X_{hjN}$, $\bar{X}_{hN} \in [\bar{X}_{hL}, \bar{X}_{hU}]$

Sample mean	$\bar{y}_{stN} = \sum_{h=1}^{L} W_{hN}\bar{y}_{hN},$ $\bar{y}_{stN} \in [\bar{y}_{stL}, \bar{y}_{stU}]$	$\bar{x}_{stN} = \sum_{h=1}^{L} W_{hN}\bar{x}_{hN},$ $\bar{x}_{stN} \in [\bar{x}_{stL}, \bar{x}_{stU}]$
Sample mean of h^{th} stratum	$\bar{y}_{hN} = \frac{1}{n_{hN}} \sum_{j=1}^{n_{hN}} y_{hjN},$ $\bar{y}_{hN} \in [\bar{y}_{hL}, \bar{y}_{hU}]$	$\bar{x}_{hN} = \frac{1}{n_{hN}} \sum_{j=1}^{n_{hN}} x_{hjN},$ $\bar{x}_{hN} \in [\bar{x}_{hL}, \bar{x}_{hU}]$
Population variance	$S_{YN}^2 = \frac{1}{N_N - 1} \sum_{h=1}^{L} \sum_{j=1}^{N_N} (Y_{hjN} - \bar{Y}_N)^2,$ $S_{YN}^2 \in [S_L^2, S_U^2]$	$S_{XN}^2 = \frac{1}{N_N - 1} \sum_{h=1}^{L} \sum_{j=1}^{N_N} (X_{hjN} - \bar{X}_N)^2,$ $S_{XN}^2 \in [S_L^2, S_U^2]$
Population variance of h^{th} stratum	$S_{yhN}^2 = \frac{1}{N_{hN} - 1} \sum_{j=1}^{N_{hN}} (Y_{hjN} - \bar{Y}_{hN})^2,$ $S_{yhN}^2 \in [S_{yhL}^2, S_{yhU}^2]$	$S_{xhN}^2 = \frac{1}{N_{hN} - 1} \sum_{j=1}^{N_{hN}} (X_{hjN} - \bar{X}_{hN})^2,$ $S_{xhN}^2 \in [S_{xhL}^2, S_{xhU}^2]$
Population coefficient of variation	$C_{yhN} = \frac{S_{yhN}}{\bar{Y}_{hN}},$ $C_{yhN} \in [C_{yhL}, C_{yhU}]$	$C_{xhN} = \frac{S_{xhN}}{\bar{X}_{hN}},$ $C_{xhN} \in [C_{xhL}, C_{xhU}]$
Population covariance	$S_{yxhN} = \frac{1}{N_h - 1} \sum_{j=1}^{N_{hN}} (y_{hjN} - \bar{Y}_{hN})(x_{hjN} - \bar{X}_{hN}),$ $S_{yxhN} \in [S_{yxhL}, S_{yxhU}]$	
Population coefficient of correlation	$\rho_{yxN} = \frac{S_{yxN}}{S_{yN} S_{xN}},$ $\rho_{yxN} \in [\rho_{yxL}, \rho_{yxU}],$	
Population coefficient of correlation of h^{th} stratum	$\rho_{yxhN} = \frac{S_{yxhN}}{S_{yhN} S_{xhN}},$ $\rho_{yxhN} \in [\rho_{yxhL}, \rho_{yxhU}]$	

The relative error terms and their expectations given below are used to derive expressions for the bias, mean squared error, and minimum mean squared error of both the existing and proposed estimators.

$$e_{yN} = \frac{\bar{y}_{stN} - \bar{Y}_N}{\bar{Y}_N}, \quad e_{yN} \in [e_{yL}, e_{yU}]$$

and

$$e_{xN} = \frac{\bar{x}_{stN} - \bar{X}_N}{\bar{X}_N}, \quad e_{xN} \in [e_{xL}, e_{xU}]$$

such that $E(e_{yN}) = E(e_{xN}) = 0$,

$$E(e_{yN}^2) = \sum_{h=1}^{L} W_h^2 f_{hN} \frac{S_{yhN}^2}{\bar{Y}_N^2} = \nu_{20N}, \quad e_{yN}^2 \in [e_{yL}^2, e_{yU}^2],$$

$$E(e_{xN}^2) = \sum_{h=1}^{L} W_h^2 f_{hN} \frac{S_{xhN}^2}{\bar{X}_N^2} = \nu_{02N}, \quad e_{xN}^2 \in [e_{xL}^2, e_{xU}^2],$$

$$E(e_{yN}e_{xN}) = \sum\nolimits_{h=1}^{L} W_h^2 f_{hN} \frac{S_{yxhN}}{\bar{Y}_N \bar{X}_N} = \nu_{11N},$$

where $f_{hN} = \left(\frac{1}{n_{hN}} - \frac{1}{N_{hN}}\right)$.

Assumptions: Before proceeding with the derivations, it is important to state the assumptions adopted in this study. In particular, small error approximations and first-order conditions are employed, as they play a vital role in statistical estimation by enabling the derivation of clear and tractable expressions for bias, variance, and mean squared error, thereby facilitating a deeper understanding of estimator behavior. Small error approximations assume that stochastic deviations in the model are sufficiently small, allowing higher-order terms to be neglected without a significant loss of accuracy. Similarly, first-order conditions are based on linear expansions around the true population parameters, capturing the dominant sources of

variation while ignoring negligible components. When applied appropriately, these assumptions enhance both the theoretical rigor and the practical interpretability of the proposed statistical procedures.

3. Existing Neutrosophic Estimators

A brief introduction to some neutrosophic estimators for the estimation of the population mean under stratified random sampling, as documented in the literature, is provided below.

3.1. Usual mean estimator

The usual mean estimator along with its variance under NSRS is given below:

$$\bar{y}_{(Nst)} = \sum_{h=1}^{L} W_{hN}\bar{y}_{hN}, \tag{3.1}$$

$$\text{var}(\bar{y}_{Nst}) = \bar{Y}_N^2 v_{20N}. \tag{3.2}$$

3.2. Ratio estimator

The traditional ratio estimator and its MSE under NSRS expressed as:

$$\bar{y}_{R(Nst)} = \bar{y}_{Nst}\left(\frac{\bar{X}_N}{\bar{x}_{Nst}}\right), \tag{3.3}$$

$$MSE(\bar{y}_{R(Nst)}) \cong \bar{Y}_N^2[v_{20N} + v_{02N} - 2v_{11N}]. \tag{3.4}$$

3.3. Product estimator

The conventional product estimator and its MSE under NSRS are given by

$$\bar{y}_{Pr(Nst)} = \bar{y}_{Nst}\left(\frac{\bar{x}_{Nst}}{\bar{X}_N}\right), \tag{3.5}$$

$$MSE(\bar{y}_{Pr(Nst)}) \cong \bar{Y}_N^2[v_{20N} + v_{02N} + 2v_{11N}]. \tag{3.6}$$

3.4. Ratio exponential estimator

The neutrosophic stratified ratio exponential estimator and its MSE are given below:

$$\bar{y}_{Re(Nst)} = \bar{y}_{Nst} \exp\left(\frac{\bar{X}_N - \bar{x}_{Nst}}{\bar{X}_N + \bar{x}_{Nst}}\right), \tag{3.7}$$

$$MSE(\bar{y}_{Re(Nst)}) \cong \bar{Y}_N^2\left[v_{20N} + \frac{1}{4}v_{02N} - v_{11N}\right]. \tag{3.8}$$

3.5. Product exponential estimator

The neutrosophic stratified product exponential estimator and its MSE are given by

$$\bar{y}_{Pe(Nst)} = \bar{y}_{Nst} \exp\left(\frac{\bar{x}_{Nst} - \bar{X}_N}{\bar{x}_{Nst} + \bar{X}_N}\right), \tag{3.9}$$

$$MSE(\bar{y}_{Pe(Nst)}) \cong \bar{Y}_N^2\left[v_{20N} + \frac{1}{4}v_{02N} + v_{11N}\right]. \tag{3.10}$$

4. Proposed Neutrosophic Estimator

Shabbir and Gupta [6] proposed an estimator for efficient estimation of the population mean under stratified sampling in the presence of classical, precise, and well-defined data. Motivated by their work and extending their framework to accommodate the uncertainty and indeterminacy inherent in neutrosophic datasets, we develop a novel neutrosophic exponential estimator to improve population mean estimation under neutrosophic stratified sampling. Exponential estimators offer greater flexibility by reducing bias when the correlation structure between variables departs from strict linearity, and their formulation enables improved performance for skewed data and heterogeneous populations. The proposed estimator is defined as follows:

$$\bar{y}_{P(Nst)} = [\lambda_{1N}\bar{y}_{Nst} + \lambda_{2N}(\bar{X}_N - \bar{x}_{Nst})]\exp\left(\frac{\bar{X}_N - \bar{x}_{Nst}}{\bar{X}_N + \bar{x}_{Nst} + 2N_N\bar{X}_N}\right), \tag{4.1}$$

where λ_{1N} and λ_{2N} represent weights, whose optimal values will be obtained in the subsequent steps by minimizing the mean squared error.

Following the assumptions and using the relative error terms defined earlier, equation (4.1) can be rewritten as

$$\bar{y}_{P(Nst)} = [\lambda_{1N}\bar{Y}_N(1+e_{yN}) + \lambda_{2N}(\bar{X}_N - \bar{X}_N(1+e_{xN}))] \times \exp\left\{-\frac{e_{xN}}{2(1+N_N)+e_{xN}}\right\},$$

$$\bar{y}_{P(Nst)} = [\lambda_{1N}\bar{Y}_N(1+e_{yN}) - \lambda_{2N}\bar{X}_N e_{xN}] \times \exp\left\{-\frac{e_{xN}}{2(1+N_N)}\left(1+\frac{e_{xN}}{2(1+N_N)}\right)^{-1}\right\}.$$

Simplifying the expression $\left(1+\frac{e_{xN}}{2(1+N)}\right)^{-1}$ with the help of the Taylor series expansion $(1+x)^{-1} = 1 - x + x^2 - x^3 + \cdots$ and retaining the expressions up to the first order of approximations, we get the following:

$$\bar{y}_{P(Nst)} \cong [\lambda_{1N}\bar{Y}_N(1+e_{yN}) - \lambda_{2N}\bar{X}_N e_{xN}] \times \exp\left(-\frac{e_{xN}}{2(1+N_N)} + \frac{e_{xN}^2}{4(1+N_N)^2}\right).$$

Again, using Taylor series expansion $\exp(x) = e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$ to expand the right hand side expression $\exp\left(-\frac{e_{xN}}{2(1+N)} + \frac{e_{xN}^2}{4(1+N)^2}\right)$, up to first order of approximation, we obtain

$$\bar{y}_{P(Nst)} \cong [\lambda_{1N}\bar{Y}_N(1+e_{yN}) - \lambda_{2N}\bar{X}_N e_{xN}] \times \left(1-\frac{e_{xN}}{2(1+N_N)} + \frac{3e_{xN}^2}{8(1+N_N)^2}\right). \quad (4.2)$$

Subtracting $\overline{Y}_N$ from both sides of equation (4.2)

$$(\bar{y}_{P(Nst)} - \overline{Y}_N) \cong \overline{Y}_N(\lambda_{1N} - 1) + \lambda_{1N}\overline{Y}_N\left(e_{yN} - \frac{e_{xN}}{2(1+N_N)} - \frac{e_{yN}e_{xN}}{2(1+N_N)} + \frac{3e_{xN}^2}{8(1+N_N)^2}\right) - \lambda_{2N}\overline{X}_N\left(e_{xN} - \frac{e_{xN}^2}{2(1+N_N)}\right). \quad (4.3)$$

Taking expectations on both sides of equation (4.3) and applying the expectation operator to the relative error terms e's, followed by substituting their respective values, we obtain

$$Bias(\bar{y}_{P(Nst)}) = E(\bar{y}_{P(Nst)} - \overline{Y}_N),$$

$$Bias(\bar{y}_{P(Nst)}) \cong \overline{Y}_N\left[(\lambda_{1N} - 1) + \lambda_{1N}\left\{\frac{3v_{02N}}{8(1+N_N)^2} - \frac{v_{11N}}{2(1+N_N)}\right\}\right] + \lambda_{2N}\overline{X}_N\frac{v_{02N}}{2(1+N_N)}. \quad (4.4)$$

Taking square on both sides of equation (4.3), we have

$$(\bar{y}_{P(Nst)} - \overline{Y}_N)^2$$

$$\cong \overline{Y}_N^2(\lambda_{1N} - 1)^2 + \lambda_{1N}^2\overline{Y}_N^2\left(2e_{yN} - \frac{e_{xN}}{(1+N_N)} + e_{yN}^2 + \frac{e_{xN}^2}{(1+N_N)^2} - \frac{2e_{yN}e_{xN}}{(1+N_N)}\right) + \lambda_{1N}\overline{Y}_N^2\left(-2e_{yN} + \frac{e_{xN}}{(1+N_N)} - \frac{3e_{xN}^2}{4(1+N_N)^2} + \frac{e_{yN}e_{xN}}{(1+N_N)}\right) + \lambda_{2N}^2\overline{X}_N^2e_{xN}^2 + \lambda_{2N}\overline{Y}_N\overline{X}_N\left(2e_{xN} - \frac{e_{xN}^2}{(1+N_N)}\right)$$

$$+ \lambda_{1N}\lambda_{2N}\overline{Y}_N\overline{X}_N\left(-2e_{xN} + \frac{2e_{xN}^2}{(1+N_N)} - 2e_{yN}e_{xN}\right). \tag{4.5}$$

Applying expectation on both sides of equation (4.5), we get the $MSE(\overline{y}_{P(Nst)})$,

$$MSE(\overline{y}_{P(Nst)}) = E(\overline{y}_{P(Nst)} - \overline{Y}_N)^2,$$

$$MSE(\overline{y}_{P(Nst)}) \cong \overline{Y}_N^2\left[(\lambda_{1N}-1)^2 + \lambda_{1N}^2\left(v_{20N} + \frac{v_{02N}}{(1+N_N)^2} - \frac{2v_{11N}}{(1+N_N)}\right)\right.$$

$$\left. - \lambda_{1N}\left(\frac{3v_{02N}}{4(1+N_N)^2} - \frac{v_{11N}}{(1+N_N)}\right)\right]$$

$$+ \lambda_{2N}^2\overline{X}_N^2 v_{02N} - \lambda_{2N}\overline{Y}_N\overline{X}_N\frac{v_{02N}}{(1+N_N)}$$

$$+ 2\lambda_{1N}\lambda_{2N}\overline{Y}_N\overline{X}_N\left(\frac{v_{02N}}{(1+N_N)} - v_{11N}\right). \tag{4.6}$$

Partially differentiating equation (4.6) with respect to λ_1 and λ_2 and equating them to zero, we obtain:

$$\frac{\partial MSE(\overline{y}_{P(Nst)})}{\partial\lambda_1} = \overline{Y}_N^2\left[2(\lambda_{1N}-1) + 2\lambda_{1N}\left(v_{20N} + \frac{v_{02N}}{(1+N_N)^2} - \frac{2v_{11N}}{(1+N_N)}\right)\right.$$

$$\left. - \left(\frac{3v_{02N}}{4(1+N_N)^2} - \frac{v_{11N}}{(1+N_N)}\right)\right]$$

$$+ 2\lambda_{2N}\overline{Y}_N\overline{X}_N\left(\frac{v_{02N}}{(1+N_N)} - v_{11N}\right) = 0, \tag{4.7}$$

$$\frac{\partial MSE(\overline{y}_{P(Nst)})}{\partial\lambda_2} = 2\lambda_{2N}\overline{X}_N^2 v_{02N} - \overline{Y}_N\overline{X}_N\frac{v_{02N}}{(1+N_N)}$$

$$+ 2\lambda_{1N}\overline{Y}_N\overline{X}_N\left(\frac{v_{02N}}{(1+N_N)} - v_{11N}\right) = 0. \tag{4.8}$$

Solving equation (4.7) and equation (4.8) simultaneously yields the optimal values of $\lambda_i \ (i = 1, 2)$ which are chosen to minimize the $MSE(\bar{y}_{P(Nst)})$,

$$\lambda_{1N}^{*} = \frac{1 - \dfrac{v_{02N}}{8(1 + N_N)^2}}{1 + v_{02N}(1 - \rho_{Nst}^2)},$$

$$\lambda_{2N}^{*} = \frac{\bar{Y}_N}{\bar{X}_N}\left[\frac{1}{2(1 + N_N)} - \lambda_{1N}^{*}\left\{\frac{1}{(1 + N_N)} - \frac{v_{11N}}{v_{02N}}\right\}\right].$$

Substituting the optimum values of $\lambda_i \ (i = 1, 2)$ in equation (4.6), we get the minimum MSE of $\bar{y}_{P(Nst)}$ which is given by

$$MSE(\bar{y}_{P(Nst)})_{\min} \cong \bar{Y}_N^2\left[\left\{1 - \frac{v_{02N}}{4(1 + N_N)^2}\right\} - \frac{\left\{1 - \dfrac{v_{02N}}{8(1 + N_N)^2}\right\}^2}{1 + v_{02N}(1 - \rho_{Nst}^2)}\right]. \quad (4.9)$$

5. Numerical Illustration

In order to substantiate the effectiveness of the proposed estimator, we employed two real-life neutrosophic data sets in a detailed numerical study. The following section provides a brief description of each data set, and their descriptive statistics are summarized in Tables 1 and 2, respectively.

Data set I

The first data set is obtained from real-world climate data for the U.S. states of Georgia and Alabama for the month of November. Selected data values are presented in Table 2 to illustrate the neutrosophic nature of dataset, characterized by interval-valued observations. For further details, see the data source: https://mrcc.purdue.edu. Among the many variables available in the dataset, the present analysis focuses on two key atmospheric indicators: relative humidity (study variable Y_{hN}) and dew point (auxiliary variable X_{hN}).

Table 2. A preview of data set I

1st stratum		2nd stratum	
Relative humidity Y_{1jN}	Dew point X_{1jN}	Relative humidity Y_{2jN}	Dew point X_{2jN}
$[\text{Min}Y_{1jL}, \text{Max}Y_{1jU}]$	$[\text{Min}X_{1jL}, \text{Max}X_{1jU}]$	$[\text{Min}Y_{2jL}, \text{Max}Y_{2jU}]$	$[\text{Min}X_{2jL}, \text{Max}Y_{2jU}]$
[17, 100]	[10, 66]	[22, 100]	[19, 68]
[12, 94]	[12, 69]	[16, 100]	[13, 66]
[16, 94]	[14, 64]	[16, 100]	[11, 64]
:	:	:	:
:	:	:	:
[28, 120]	[20, 80]	[38, 55]	[23, 31]

Table 3. Summary statistics of data set I

Measure	**Parameter**	**1st stratum**	**2nd stratum**
Population size	N_N	[19, 19]	[22, 22]
Sample size	n_N	[7, 7]	[8, 8]
Mean of study variable	$[\bar{Y}_{NL}, \bar{Y}_{NU}]$	[28.21, 96.47]	[31.77, 93.86]
Mean of auxiliary variable	$[\bar{X}_{NL}, \bar{X}_{NU}]$	[19.58, 61.95]	[22.55, 62.23]
Coefficient of variation of study variable	$[C_{yNL}, C_{yNU}]$	[0.141, 0.826]	[0.126, 0.689]
Coefficient of variation of auxiliary variable	$[C_{xNL}, C_{xNU}]$	[0.230, 0.555]	[0.139, 0.579]
Standard deviation of study variable	$[S_{yNL}, S_{yNU}]$	[13.55, 23.30]	[11.78, 21.88]
Standard deviation of auxiliary variable	$[S_{xNL}, S_{xNU}]$	[10.86, 14.22]	[8.670, 13.06]
Coefficient of correlation	$[\rho_{xyNL}, \rho_{xyNU}]$	[0.946, 0.941]	[0.885, 0.948]

Data set II

The second dataset consists of monthly temperature and dew point observations for Lahore, Punjab, Pakistan, during the year 2023. This dataset was selected because the values of both variables are not fixed; rather, they fluctuate within uncertain intervals, making it well suited for neutrosophic analysis. In this study, the four seasons: Winter, Spring, Summer, and

Autumn are considered as strata. Temperature is treated as the study variable Y_{hN} while dew point serves as the auxiliary variable X_{hN}. The data were obtained from the weather website (https://www.wunderground.com/history/daily/OPLH/ date/2005-1-3).

Table 4. Summary statistics of data set II

Parameters	1st stratum (winter)	2nd stratum (spring)	3rd stratum (summer)	4th stratum (Autumn)
N_N	[90, 90]	[92, 92]	[92, 92]	[91, 91]
n_N	[20, 20]	[20, 20]	[20, 20]	[20, 20]
$[\bar{X}_{NL}, \bar{X}_{NU}]$	[42.300, 51.278]	[52.923, 62.272]	[71.326, 77.880]	[61.319, 67.615]
$[\bar{Y}_{NL}, \bar{Y}_{NU}]$	[47.487, 68.789]	[67.011, 89.456]	[79.663, 95.826]	[67.519, 86.000]
$[C_{yNL}, C_{yNU}]$	[0.122, 0.126]	[0.106, 0.107]	[0.061, 0.064]	[0.092, 0.135]
$[C_{xNL}, C_{xNU}]$	[0.095, 0.133]	[0.084, 0.085]	[0.051, 0.065]	[0.116, 0.131]
$[S_{yNL}, S_{yNU}]$	[5.181, 8.381]	[7.005, 9.537]	[5.088, 5.864]	[7.918, 9.117]
$[S_{xNL}, S_{xNU}]$	[4.846, 5.642]	[4.495, 5.294]	[3.947, 4.619]	[7.839, 8.032]
$[\rho_{xyNL}, \rho_{xyNU}]$	[0.456, 0.661]	[0.298, 0.565]	[0.089, 0.523]	[0.770, 0.921]

Performance indicators

To evaluate the superiority of the proposed estimator over the existing ones, two performance indicators are employed: Mean Square Error (MSE) and Percentage Relative Efficiency (PRE). MSE measures the divergence of the estimator's values from the true parameter value. The MSEs of the existing estimators $\bar{y}_{Nst}$, $\bar{y}_{R(Nst)}$, $\bar{y}_{Pr(Nst)}$, $\bar{y}_{Re(Nst)}$ and $\bar{y}_{Pe(Nst)}$ are computed using equations (3.2), (3.4), (3.6), (3.8) and (3.10), respectively. The MSE of the proposed estimator $\bar{y}_{P(Nst)}$ is determined using equation (4.9).

PRE quantifies the improvement achieved by the proposed estimator. It is a comparative measure defined as the ratio of the MSE of an existing estimator to that of the proposed estimator, using the usual estimator $\bar{y}_{Nst}$ as

the baseline. PRE is computed as

$$PRE(\bar{y}_*) = \frac{MSE(\bar{y}_{Nst})}{MSE(\bar{y}_*)} \times 100,$$

where

$$(*) = \bar{y}_{R(Nst)}, \bar{y}_{Pr(Nst)}, \bar{y}_{Re(Nst)}, \bar{y}_{Pe(Nst)}, \bar{y}_{P(Nst)}.$$

Tables 5 and 6 present the comparative performance of the estimators, expressed in terms of indeterminacy intervals of $[MSE_L, MSE_U]$ and $[PRE_L, PRE_U]$ for data sets I and II, respectively. A narrower interval with lower MSE values and a wider interval with higher PRE values indicate greater efficiency relative to competing estimators.

It is evident from Table 5 that, for data set I, the proposed estimator $\bar{y}_{P(Nst)}$ exhibits the lowest MSE interval [1.199, 2.588] and the widest PRE interval [559.774, 845.215]. A similar pattern is observed for data set II in Table 6, where the proposed estimator attains an MSE interval [0.310, 0.573] and a PRE interval [117.630, 164.592].

These results demonstrate that the proposed estimator $\bar{y}_{P(Nst)}$ consistently outperforms its counterparts across all datasets.

Table 5. MSEs and PREs of existing and proposed estimators for data set I

Estimator	$[MSE_L, MSE_U]$	$[PRE_L, PRE_U]$
$\bar{y}_{Nst}$	[6.716, 21.873]	[100, 100]
$\bar{y}_{R(Nst)}$	[2.863, 3.358]	[234.529, 651.289]
$\bar{y}_{Pr(Nst)}$	[37.045, 65.127]	[18.129, 33.585]
$\bar{y}_{Re(Nst)}$	[1.480, 9.523]	[229.679, 453.725]
$\bar{y}_{Pe(Nst)}$	[18.571, 40.407]	[36.164, 54.131]
$\bar{y}_{P(Nst)}$	**[1.199, 2.588]**	**[559.774, 845.215]**

Table 6. MSEs and PREs of existing and proposed estimators for data set II

Estimator	$[MSE_L, MSE_U]$	$[PRE_L, PRE_U]$
$\bar{y}_{Nst}$	[0.511, 0.675]	[100, 100]
$\bar{y}_{R(Nst)}$	[0.371, 0.762]	[88.567, 137.785]
$\bar{y}_{Pr(Nst)}$	[1.613, 1.719]	[31.679, 39.278]
$\bar{y}_{Re(Nst)}$	[0.321, 0.577]	[116.937, 159.342]
$\bar{y}_{Pe(Nst)}$	[0.941, 1.056]	[54.259, 63.955]
$\bar{y}_{P(Nst)}$	**[0.310, 0.573]**	**[117.630, 164.592]**

To facilitate a clear comparative evaluation of MSEs and PREs, bar charts are employed, with lower bounds shown on the left and upper bounds on the right. Figures 1 and 3 present the MSEs for data sets I and II, respectively. These figures clearly indicate that the proposed estimator $\bar{y}_{P(Nst)}$ achieves the smallest bar lengths under both bounds compared to all competing estimators. Similarly, Figures 2 and 4 display the PREs for datasets I and II, respectively, where the proposed estimator consistently attains higher bar lengths under both bounds than the competing estimators $\bar{y}_{Nst}$, $\bar{y}_{R(Nst)}$, $\bar{y}_{Pr(Nst)}$, $\bar{y}_{Re(Nst)}$ and $\bar{y}_{Pe(Nst)}$. Overall, Figures 1-4 clearly demonstrate that $\bar{y}_{P(Nst)}$ is the most efficient estimator under both lower and upper bounds.

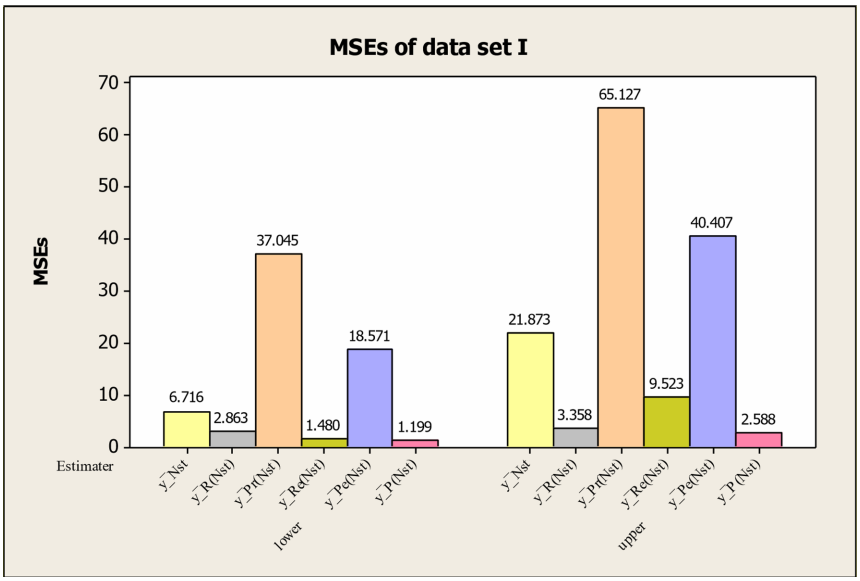

Figure 1. MSEs of data set I.

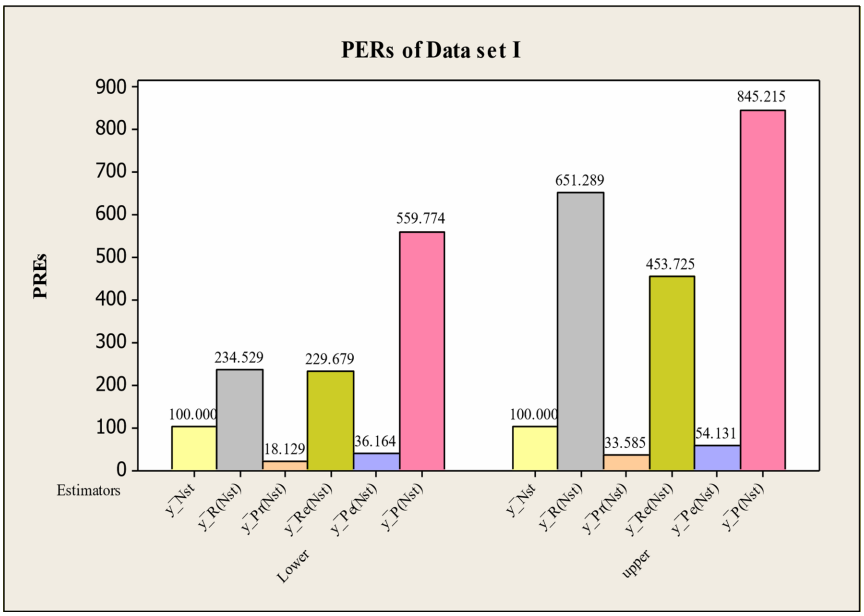

Figure 2. PREs of data set I.

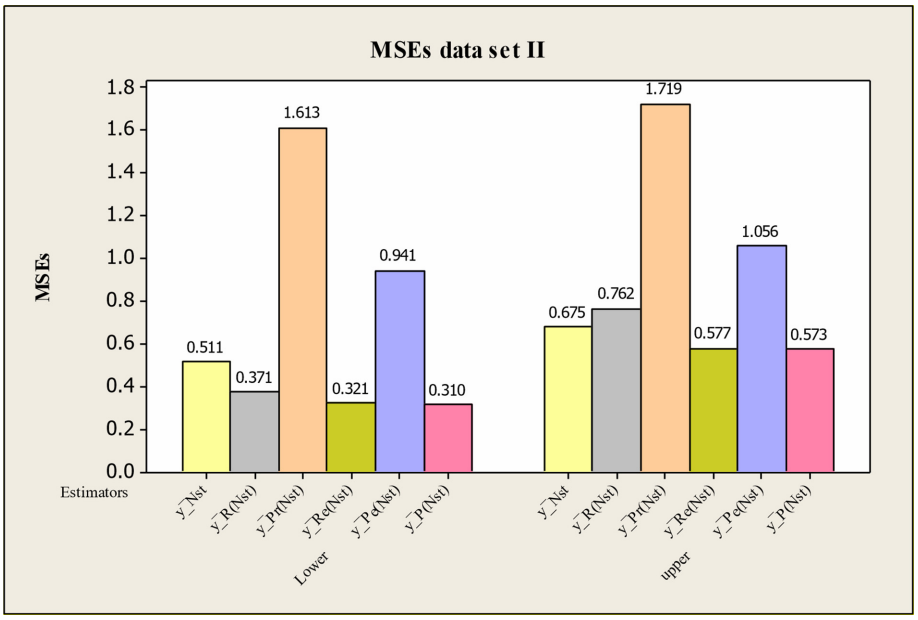

Figure 3. MSEs of data set II.

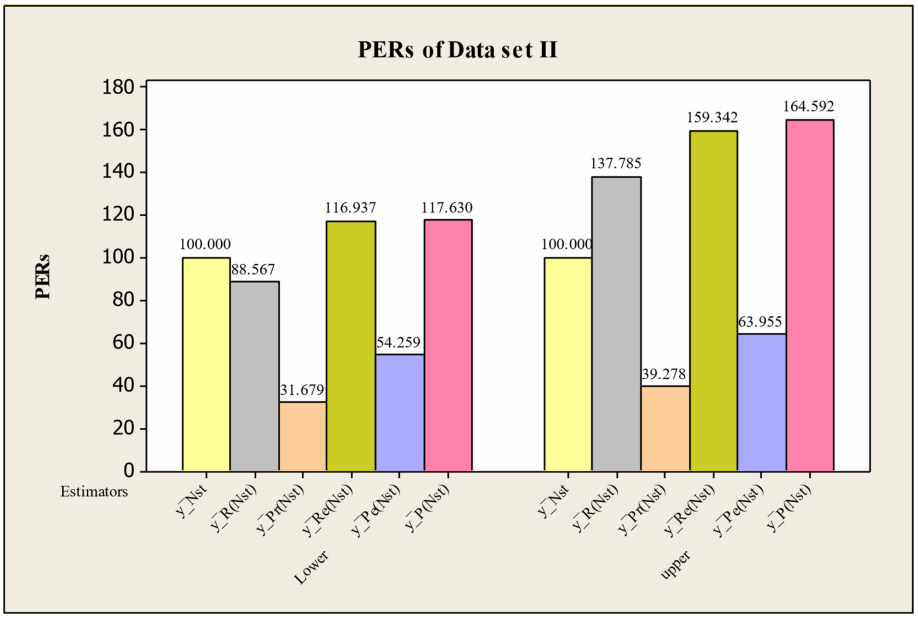

Figure 4. PREs of data set II.

Furthermore, the reduction in error achieved by the proposed estimator, relative to the existing estimators, is assessed using MSEs. The percentage reduction in error is computed as

$$\text{Percentage reduction in error} = \frac{MSE(\bar{y}_{**}) - MSE(\bar{y}_{P(Nst)})}{MSE(\bar{y}_{**})} \times 100,$$

where

$$(**) = \bar{y}_{Nst},\ \bar{y}_{R(Nst)},\ \bar{y}_{Pr(Nst)},\ \bar{y}_{Re(Nst)},\ \bar{y}_{Pe(Nst)}.$$

Alternatively, Percentage Relative Efficiency (PRE) may be used in place of MSE in the above formulation to assess the reduction in error.

For data set I, the proposed estimator $\bar{y}_{Nst}$ achieves percentage reduction in error of [97%, 96%], [93%, 94%], [82%, 88%], [58%, 23%] and [19%, 72%] relative to $\bar{y}_{Pr(Nst)}$, $\bar{y}_{Pe(Nst)}$, $\bar{y}_{Nst}$, $\bar{y}_{R(Nst)}$, $\bar{y}_{Re(Nst)}$, respectively. The highest reduction in error is observed with respect to $\bar{y}_{Pr(Nst)}$, while the lowest reduction is achieved relative to $\bar{y}_{Re(Nst)}$.

Similarly, for data set II, the proposed estimator $\bar{y}_{Nst}$ yields error reduction of [81%, 67%], [67%, 46%], [39%, 15%], [17%, 25%] and [3%, 0.6%] when compared with $\bar{y}_{Pr(Nst)}$, $\bar{y}_{Pe(Nst)}$, $\bar{y}_{Nst}$, $\bar{y}_{R(Nst)}$, $\bar{y}_{Re(Nst)}$, respectively. As in data set I, the maximum reduction in error is observed relative to $\bar{y}_{Pr(Nst)}$ whereas the minimum reduction is observed relative to $\bar{y}_{Re(Nst)}$.

In line with this context, despite the comprehensive analysis presented in this study, several limitations merit consideration. The methodology primarily relies on first-order approximations, which, while facilitating tractable analytical expressions, may not fully account for higher-order effects. Moreover, the investigation is conducted using small stratified samples, potentially limiting the generalizability of the findings to broader or more heterogeneous populations. Additionally, the study does not explore optimal allocation strategies within stratified sampling, which could further

enhance estimator efficiency. Acknowledging these limitations not only ensures transparency and scientific rigor but also provides valuable guidance for future research aimed at extending and refining the proposed methodology.

6. Conclusion

This study proposes a neutrosophic stratified exponential ratio-type estimator to enhance estimation of population mean in heterogeneous neutrosophic populations, effectively addressing the inherent uncertainty and imprecision in such data. A mathematical framework is developed to derive expressions for the bias and Mean Squared Error (MSE) using first-order approximations. Both theoretical and empirical analyses demonstrate the superior performance of the proposed estimator, as it consistently yields lower MSEs and higher percent relative efficiencies (PREs) than conventional estimators. To further support these findings, graphical illustrations in the form of bar charts are included to visually compare the MSEs and PREs under lower and upper bounds. The proposed estimator outperforms the neutrosophic stratified unbiased estimator, separate ratio and product estimators, as well as existing exponential estimators. Overall, these results underscore its practical effectiveness and recommend the proposed estimator as a reliable and efficient tool for neutrosophic stratified sampling in environments characterized by indeterminate data.

Acknowledgment

The authors would like to express their gratitude to the editors and reviewers for their wise suggestions on how to improve the work.

References

[1] A. Ali H. Ahmadini, R. Varshney, Mradula and I. Ali, On multivariate-multiobjective stratified sampling design under probabilistic environment: A fuzzy programming technique, Journal of King Saud University - Science 33(5) (2021), 101448. https://doi.org/10.1016/j.jksus.2021.101448

[2] C. Kadilar and H. Cingi, A new ratio estimator in stratified random sampling, Communications in Statistics - Theory and Methods 34(3) (2005), 597-602. https://doi.org/10.1081/sta-200052156

[3] N. Koyuncu and C. Kadilar, On improvement in estimating population mean in stratified random sampling, Journal of Applied Statistics 37(6) (2010), 999-1013. https://doi.org/10.1080/02664760903002675

[4] S. Kumar, S. P. Kour, M. Choudhary and V. Sharma, Determination of population mean using neutrosophic, exponential-type estimator, Lobachevskii Journal of Mathematics 43(11) (2022), 3359-3367. https://doi.org/10.1134/s1995080222140219

[5] G. S. Rao and M. Aslam, Inspection plan for COVID-19 patients for Weibull distribution using repetitive sampling under indeterminacy, BMC Medical Research Methodology 21(1) (2021). https://doi.org/10.1186/s12874-021-01387-7

[6] J. Shabbir and S. Gupta, On estimating finite population mean in simple and stratified random sampling, Communications in Statistics - Theory and Methods 40(2) (2010), 199-212. https://doi.org/10.1080/03610920903411259

[7] R. Sharma and S. Malik, Exponential type ratio and product estimator of finite population mean in stratified Sampling, Indian Journal of Science and Technology 17(40) (2024), 4168-4176. https://doi.org/10.17485/ijst/v17i40.2237

[8] A. Singh, H. Kulkarni, F. Smarandache and G. K. Vishwakarma, Computation of separate ratio and regression estimator under Neutrosophic stratified sampling: an application to climate data, Journal of Fuzzy Extension and Applications 5(4) (2024a), 605-621.

[9] A. Singh, V. G. Patkar, H. Kulkarni and G. K. Vishwakarma, Computation of separate product and ratio exponential estimators under neutrosophic stratified sampling: application to climate data, Research Square, 2024b.

[10] Z. Tahir, H. Khan, M. Aslam, J. Shabbir, Y. Mahmood and F. Smarandache, Neutrosophic ratio-type estimators for estimating the population mean, Complex & Intelligent Systems 7(6) (2021), 2991-3001. https://doi.org/10.1007/s40747-021-00439-1

[11] N. D. Thanh, L. H. Son and M. Ali, Neutrosophic recommender system for medical diagnosis based on algebraic similarity measure and clustering, 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2017, pp. 1-6.

[12] A. Ullah, J. Shabbir, A. M. Alomair and F. K. Alarfaj, Compromise optimum allocation in neutrosophic multi-character survey under stratified random sampling using neutrosophic fuzzy programming, Heliyon 10(7) (2024), e28327. https://doi.org/10.1016/j.heliyon.2024.e28327